

# Biochemical pathways simulation

A. Cangiani

August 22, 2012

## Abstract

In this short note we review deterministic simulation of biochemical pathways, i.e. networks of biochemical reactions obeying the law of mass action. It is meant as a basis for the MATLAB code, written by the author, which permits easy input and simulation of general biochemical networks. This work was carried out for the European Project ‘CardioWorkBench’.

## 1 Introduction

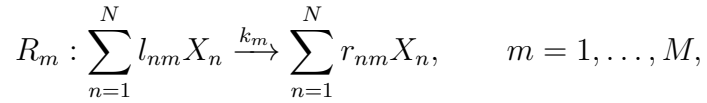
These notes give a short guidance to deterministic simulation of biochemical pathways. We follow the approach introduced by Ullah *et al.* [5], basis of the system biology sbtoolbox [4] freely available from [1]. For a comprehensive list of system biology software see [2].

The aim is to easily input and simulate a general biochemical network of reactions obeying the law of mass action. The pathway specifications (rates, reactants, and products of each reaction) have to be easily and flexibly importable. And the evaluation of the resulting system of ODEs expressing the model’s dynamics have to be computationally efficient. We optimise previous implementations and show how to efficiently calculate the ODEs jacobian. We also mention how to easily include parameters like phases and volume scalings. A user-friendly MATLAB code implementing the algorithm discussed in these notes can be downloaded from [3].

## 2 Modeling

We consider simulating a biochemical reaction pathway (network) assuming that it can be decomposed into unidirectional elementary reactions and that the law of mass action can be applied to each reaction.

Let  $\mathbf{X} = (X_n)_{N \times 1}$  represent the molecular species,  $\mathbf{R} = (R_m)_{1 \times M}$ , the elementary reactions,  $\mathbf{k} = (k_m)_{1 \times M}$  the *rate* coefficients. The set of elementary reactions forming the pathway can be written as



where the reactions' *stoichiometric* coefficients  $l_{nm}$  and  $r_{nm}$  are non-negative integers. The step change in the number of molecules  $X_n$  due to reaction  $R_m$  is given by

$$d_{nm} = r_{nm} - l_{nm}.$$

We collect these coefficients into the input and output stoichiometric matrices  $\mathbf{L} = (l_{nm})_{N \times M}$  and  $\mathbf{R} = (r_{nm})_{N \times M}$ , and the step change matrix  $\mathbf{D} = (d_{nm})_{N \times M}$ .

In practice, it is convenient to input the information contained in the matrices  $\mathbf{L}$  and  $\mathbf{R}$  by storing their (few) non-zero entries values and indices, as it is done in any matrix *sparse* representation. This process is described below with an example.

The law of mass action implies that the dynamics of the molar concentrations  $x_n = [X_n]$ ,  $n = 1, \dots, N$  is described by a system of ODEs, each called a *rate equation*, that we can easily write down in terms of the reactions' coefficients. The set of molar concentrations  $\mathbf{x} = (x_n)_{N \times 1}$  must satisfy the system of first order ODEs

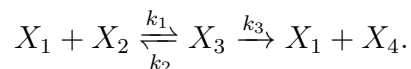
$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) \quad \text{with} \quad \mathbf{f} = (f_n)_{N \times 1} \quad \text{and} \quad (1)$$

$$f_n(\mathbf{x}) = \sum_{m=1}^M d_{nm} \left( k_m \prod_{i=1}^N x_i^{l_{im}} \right). \quad (2)$$

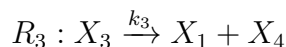
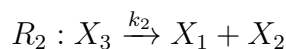
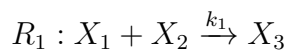
The parameter  $d_{nm}$  is different from zero if the  $n$ -th variable is involved in the  $m$ -th reaction, the rate of which is given by the expression in parentheses. Note that all this can be generalized to include more complex dynamics

by allowing the stoichiometric coefficients to be non-negative real numbers (Generalized Mass Action), see [6].

**Example.** Consider the basic enzyme-kinetic reaction



We decompose the reaction pathway into the elementary reactions



To this sequence of reactions we associate the following coefficient matrices used to compute (2):

$$D = \begin{bmatrix} -1 & 1 & 1 \\ -1 & 1 & 0 \\ 1 & -1 & -1 \\ 0 & 0 & 1 \end{bmatrix}, \quad L = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

These matrices can be obtained from a compact representation, that facilitates the input of the model data. For instance, the matrix  $L$  specifying the reactants, can be computed from

$$L^i = \begin{bmatrix} 1 & 3 & 3 \\ 2 & 0 & 0 \end{bmatrix}, \quad L^v = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

The first row of  $L^i$  (the superscript “i” stands for *index*) tells us that the first reaction involves the first and second reactant, and so on. The matrix  $L^v$  (“v” for *value*) collects the related non-zero stoichiometric coefficients  $l_{nm}$ . All remaining entries are filled with zeros. Similarly, we decompose the matrix  $R$  which contains the information about the reactions’ products into two matrices  $R^i$  and  $R^v$ .

In realistic biochemical pathways, there are many reactants and reactions involved but few reactants acting on the single elementary reaction. Consequently, most of the entries of the matrices  $L$ ,  $R$  (and  $D$ ) are zero. The compact (sparse) representation above represents a good compromise between user friendliness and computational efficiency.

### 3 Implementation and extensions

The process of simulating a biochemical pathway can be decomposed into two tasks: the first is to set up the ODE system by obtaining the matrices of stoichiometric coefficients  $\mathbf{L}$  and  $\mathbf{R}$ , and the vector of reaction rates  $\mathbf{k}$ ; the second is the actual numerical solution (time-stepping) of the system with the given concentrations initial values. The computational cost of time-stepping is dominated by the evaluation of the ODEs' function, thus it is crucial that this is done efficiently.

The implementation of  $\mathbf{f}(\mathbf{x})$  presented in [5] uses the matrices  $\mathbf{L}$  and  $\mathbf{D}$ . Letting  $\mathbf{xx} = (\mathbf{x}, \dots, \mathbf{x})_{N \times M}$  be a matrix whose  $M$ -columns are copies<sup>1</sup> of  $\mathbf{x}$ , the expression in (2) can be evaluated as follows:

$$\mathbf{f}(\mathbf{x}) = \mathbf{D} * (\mathbf{k} .* \text{prod}(\mathbf{xx} .^{\wedge} \mathbf{L}, 1))',$$

where  $*$  denotes matrix multiplication,  $.*$  and  $.^{\wedge}$  entry-wise matrix multiplication and exponentiation,  $\text{prod}(\cdot, 1)$  multiplication along columns, and  $'$  transposition (these notational conventions are in accordance with MATLAB's syntax).

In [5], it is also proposed a code that speeds-up the evaluation of  $\mathbf{f}(\mathbf{x})$  by avoiding multiplications by zero entries. This can be achieved more simply by using the compact representation above, and, if sparse matrices representation is implemented, by creating  $\mathbf{D}$  as a sparse matrix. A further straightforward computational optimization is obtained by multiplying each column of  $\mathbf{D}$  with the corresponding rate constant in  $\mathbf{k}$ . Denoting the result of such multiplication by  $\mathbf{Dk}$ , the computation of  $\mathbf{f}(\mathbf{x})$  is reduced to the following steps:

1. using  $\mathbf{L}^i$ , form the matrix  $\mathbf{X} = \begin{bmatrix} x_1 & x_3 & x_3 \\ x_2 & 0 & 0 \end{bmatrix};$
2. evaluate  $\mathbf{f}(\mathbf{x})$  as:  $\mathbf{f}(\mathbf{x}) = \mathbf{Dk} * \text{prod}(\mathbf{X} .^{\wedge} \mathbf{L}^v, 1)',$

with the convention that the empty product  $0^0 = 1$ .

Biochemical models often include parameters like, for instance, phases and volume scalings. It is useful to implement such parameters separately from the stoichiometric coefficients. This is easily done by multiplying the appropriate entry of  $\mathbf{Dk}$  by the given parameter.

---

<sup>1</sup>in MATLAB, this is achieved by the command `xx= repmat(x,1,M)`

As the ODE system generated by biochemical pathways is non-linear, it is also useful to have a routine that evaluates the jacobian  $J_{\mathbf{f}}(\mathbf{x})$  of the ODE system function  $\mathbf{f}$ . The entries of the jacobian are given by:

$$\frac{\partial f_n}{\partial x_j}(\mathbf{x}) = \sum_{m=1}^M dk_{nm} \left( l_{jm} x_j^{\max(l_{jm}-1, 0)} \prod_{i \neq j} x_i^{l_{im}} \right), \quad j, n = 1, \dots, N. \quad (3)$$

We proceed with the evaluation of the jacobian processing it by columns, i.e. evaluating  $\frac{\partial \mathbf{f}}{\partial x_j}$  for  $j = 1, \dots, N$ .

Let  $j \in \{1, \dots, N\}$  be given. The evaluation of (3) is simplified by limiting the summation to the reactions involving  $x_j$ , and by limiting the multiplications to the reactants involved in the given reaction. This latter simplification is already embedded in our compact representation. As for the summation, we define reduced matrices  $\mathbf{Dk}_{x_j}$ ,  $\mathbf{X}_{x_j}$ , and  $\mathbf{L}_{x_j}^v$  obtained from the corresponding matrices  $\mathbf{Dk}$ ,  $\mathbf{X}$ , and  $\mathbf{L}^v$  by considering only those columns associated to reactions that have  $x_j$  among the input reactants.

We need to derive the entries of  $\mathbf{X}_{x_j} \wedge \mathbf{L}_{x_j}^v$  with respect to the variable  $x_j$ . To this end, we define a new row-vector  $\mathbf{l}_{x_j}$  collecting the values  $l_{jm}$  with  $m$  such that  $x_j$  is a reactant of reaction  $m$ , and a new matrix  $\tilde{\mathbf{L}}_{x_j}^v$  obtained from  $\mathbf{L}_{x_j}^v$  by replacing the entries indexed  $jm$  with  $l_{jm} - 1$ . Notice that all such matrix manipulations are made particularly easy by MATLAB built-in vector/matrix manipulation tools. In this way, we can express (and calculate) the  $j$ -th column of the jacobian matrix in compact form as:

$$\frac{\partial \mathbf{f}}{\partial x_j}(\mathbf{x}) = \mathbf{Dk}_{x_j} * \left( \mathbf{l}_{x_j} .* \text{prod}(\mathbf{X}_{x_j} \wedge \tilde{\mathbf{L}}_{x_j}^v) \right)'.$$

Finally, let us discuss the problem of numerically solving (1). The main characteristics of such system of ODEs are the following. Unless all elementary reactions are of 0-th or 1-st order, the system is nonlinear. Moreover, the system is generally *stiff*. Thus, it is compulsory to consider variable time-stepping and employ stiff ODE solvers. Standard ODE solvers packages include robust stiff solvers (for instance, MATLAB's ode15s), that are fast enough if only a few simulations are needed. If, on the other hand, a large number of simulations is required as in parameter estimation, then it may be preferable to code *ad hoc* solvers that take into account the peculiarities of the ODE systems generated by biochemical pathways.

## References

- [1] <http://www.sbtoolbox.org>.
- [2] <http://www.sbml.org>.
- [3] <http://www2.le.ac.uk/departments/mathematics/extranet/staff-material/staff-profiles/ac433>.
- [4] H. Schmidt and M. Jirstrand. Systems biology toolbox for matlab: A computational platform for research in systems biology. *Bioinformatics*, 22(4):514–515, 2006.
- [5] M. Ullah, H. Schmidt, K. H. Cho, and O. Wolkenhauer. Deterministic modelling and stochastic simulation of biochemical pathways using matlab. *Syst Biol (Stevenage)*, 153(2):53–60, March 2006.
- [6] E. O. Voit. *Computational Analysis of Biochemical Systems : A Practical Guide for Biochemists and Molecular Biologists*. Cambridge University Press, September 2000.